International Journal of Scientific & Engineering Research, Volume 8, Issue 3, March-2017
ISSN 2229-5518

Speaker recognition using pattern recognition neural network and feedforward neural network

Neha Chauhan
Birla Institute of Technology, Mesra, Ranchi

Abstract— Speaker Recognition is the computing task of validating a user's claimed identity using speech characteristics. Main objective of speech recognition system is to communication with a device through our voice. Mel frequency Cepstral Coefficient (MFCC) features are combined with pitch and root mean square values and tested for improvement in efficiency. Feed forward artificial neural network and pattern recognition neural network is used as a classifier. System was tested for 30 speakers. Better efficiency is obtained when MFCC features are combined with pitch and rms value over single MFCC features. Feed forward neural network gives better efficiency over pattern recognition neural network.70% of total samples are used for training and 30% of total samples are used for system testing. Simulation is done using matlab software. Accuracy of system is calculated using confusion matrix.

.

Index Terms— ANN, MFCC, ASR, RMS, PITCH, PATTERN RECOGNITION, FEED FORWARD NEURAL NETWORK,

1 INTRODUCTION

Speaker identification is the process of determining which registered speaker provides a given utterance on the basis his/her recorded voice samples[1].The main aim of speaker identification is to identify a speaker from the set of different speakers on the basis of his/her voice samples. Speaker recognition system is mainly divided into three parts: Open set vs. close set, Text independent vs. text dependent and speaker identification and speaker verification. Text independent speaker identification system is implemented for close set system.

During Text independent speaker identification speaker can utter anything; his utterance does not depend on content of speech. Close set speaker identification system is designed for fixed number of speakers [2].

MFCC are widely used as feature extraction technique and can be obtained by taking Fourier transform of windowed speech signal [3].

Speaker recognition system helps in the basic purpose of speaker identification which forms a formidable domain in the field of speaker recognition. The system designed has potential in several security applications can be used in different access control.

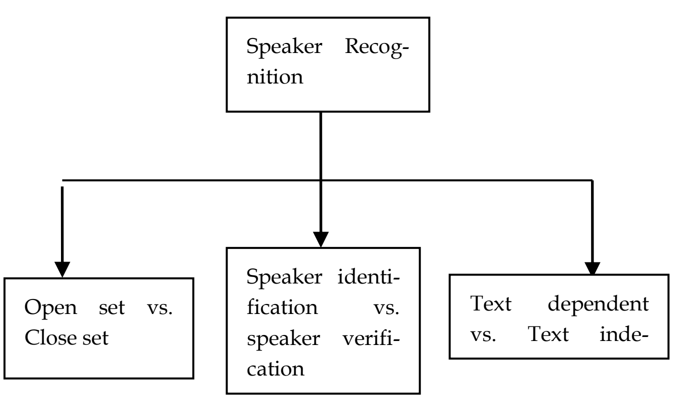

Fig.1 Speaker Recognition system

Paper is divided into four sections: Database preparation, feature extraction, classification, result and conclusion and future scope.

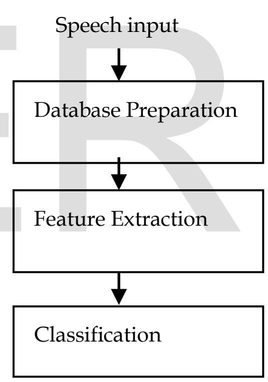

Fig 2.Research Methodology

II.DATABASE PREPARATION

Voice of different speakers are recorded on wave pad software at sampling frequency 8Khz. Database of 30 speakers are recorded and each speaker has 32 samples. All 32 samples of each speaker are segmented manually using wavepad software. Feature of each speaker is calculated using different feature extraction techniques for each recorded database [4].Total 960 samples are recorded and feature of each sample is calculated using MFCC, centroid and spectral subtraction methods.

III.FEATURE EXTRACTION

Feature Extraction is a process of extracting useful information of speech samples and discarding unwanted information like noise [5]. MFCC is mostly used feature extraction techniques. Other Feature Extraction techniques used are pitch and rms.Combination of different features is used to build the model. Due to Non stationary nature of speech signal it is better to perform feature extraction which reduce the variability of speech signals.

International Journal of Scientific & Engineering Research, Volume 8, Issue 3, March-2017
ISSN 2229-5518

A. Pitch

Pitch is a quality of sound that measures degree of highness and lowness of tone. Major difference between male and female voice is pitch. Female voice has comparatively high pitch as compare to male voice.MFCC value gives information of vocal tract while pitch represents voice source features. Pitch represents the perceived fundamental frequency (F0) of a sound and is one of the major auditory attributes of sounds along with loudness and quality [6].

B.RMS

Root mean square value is used to calculate size of the signal. Size of the signal is related to the strength of the signal. RMS value of a signal is calculated as the square root of average of squared value of the signal.RMS value of speech signal is calculated using matlab.Every speech sample generates single rms value [6].

B.Mel Frequency cepstral coefficient

MFCC is an audio feature extraction technique which extract features from speech samples similar to the one used by humans for hearing speech [7].

MFCC is one of the most popular feature extraction technique used to extract the important feature of speech signal discard all the unwanted information. MFCC is prone to noise which reduces system efficiency and therefore when MFCC values are combined with spectral features increases the efficiency of system. During MFCC speech signals are first divided into small frames consisting of arbitrary number of samples. Overlapping of the frame is done to preserve the smallest unit of sound which are phonemes. Hamming window is applied to each frame for the smooth transition. Filter coefficient of hamming window is calculated using formula (n) is window function

$$W(n)=.54-.46\cos(2\pi n/N-1), 0\leq n\leq N-1 \tag{1}$$

Where N is total number of sample and n is current sample.

Fast Fourier transform of each frame is calculated which speed up the process. Logarithmic Mel scale is applied to FFT frame which is linear up to 1 KHz and logarithmic at greater frequencies. The relationship between frequency of speech and Mel scale can be established as:

$$m_f = 2595\log_{10}(1+f/700) \tag{2}$$

Last step is to calculate discrete cosine transform which decorrelates the features and arranges them in descending order of information, they contain about speech signal. Hence first 13 coefficients are used as MFCC features for creating model. [7][8].MFCC, spectral centroid and spectral subtraction values are used to train neural network and efficiency is calculated. MFCC steps are shown in fig2.

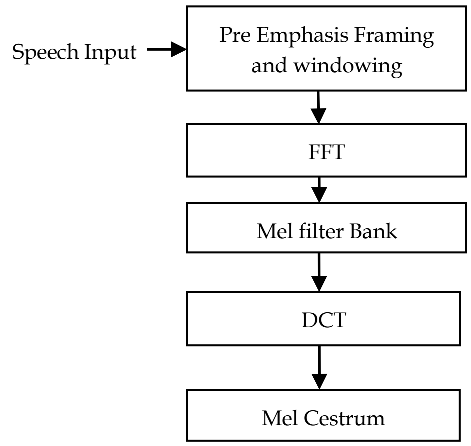

Fig 3. Block Diagram of MFCC

IV. CLASSIFICATION

Classification is the process of identifying unknown speaker by matching his/her feature with existing database using classifier. Feed forward neural network and pattern recognition neural network are used as classifier.

Features of 30 speakers are taken as an input. Classification is mainly done in two stages which are Training and testing. During Training phase system learns to perform the task. Accuracy of system is tested during testing phase.

A. Feed forward and back propagation neural network.

Feed forward neural network mainly consist of three layers: input layer, hidden layer and output layer. Direction of the neurons is forward, no loop is formed. Feed forward neural network is a part of back propagation network. Main aim of the back propagation network is reduce the error and to obtain the desired output value. Error is calculated by subtracting the actual output value from the desired output value [8].

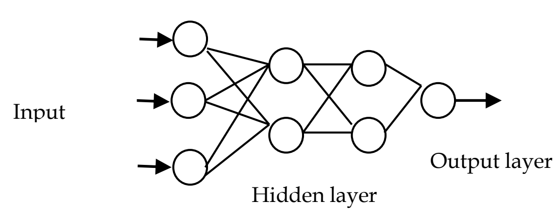

Fig.4 Feed forward neural network

B.Pattern Recognition neural network

Pattern recognition is one of the important steps in Image Processing and speech processing. First step in the pattern recognition is to select a set of features or attributes from the speech sample that will be used to classify the pattern. Next, the original pattern must be transformed into a representation that can be easily manipulated programmatically [9]. In case speaker identification feed forward neural network performs better than pattern recognition.

V.RESULT AND DISCUSSION

International Journal of Scientific & Engineering Research, Volume 8, Issue 3, March-2017
ISSN 2229-5518

A. Speaker identification result with feed forward neural network

TABLE1. AVERAGE SPEAKER IDENTIFICATION EFFICIENCY USING FFNN

Number of speakers	*%Ac-curacy using MFCC*	*%Accuracy using-MFCC+RMS*	*%Accuracy using MFCC+PITCH*	*%Accuracy using MFCC+RMS +PITCH*
10	84.7	87	86.3	87.5
20	75.2	78.4	76.9	84.1
30	50	64.7	60	65

B. Speaker identification result with pattern recognition neural network

TABLE2. AVERAGE SPEAKER IDENTIFICATION EFFICIENCY USING PATTERN RECOGNITION

Number of speakers	*%Ac-curacy using MFCC*	*%Accuracy using-MFCC+RMS*	*%Accuracy using MFCC+PITCH*	*%Accuracy using MFCC+RMS +PITCH*
10	78.4	76.3	76.9	81.9
20	66	82.2	70.9	76
30	50	64.1	50	64

Above table shows the efficiency of different number of speakers with different feature extraction techniques using FFNN and pattern recognition neural network. From above table we observe that highest efficiency is obtained using combination of MFCC, pitch and rms. FFNN gives better result for speaker identification as compared to pattern recognition technique.

Results of speaker identification system shows that efficiency of system increases when MFCC is combined with other features. Combination of MFCC.pitch and rms gives better result over other combinations. Also overall accuracy of speaker identification system using FFNN is better in comparison with pattern recognition techniques.

VI.CONCLUSION AND FUTURE SCOPE

In this paper, performance evaluation of various combinations of features is successfully studied for speaker identification system using feed forward neural network and pattern recognition neural network.

Initially, the system was tested with only MFCC features and a low efficiency of 50% was found for speaker identification system.To improve the efficiency of the system, we experimented on adding more information to our training matrix by appending features of pitch and rms. This resulted in improvement of efficiency to 10-15 % .Result was tested for 30 speakers. System show best result when MFCC features are combined with pitch and rms value. Also comparison between FFNN and pattern recognition is done and it is observed FFNN give better result over pattern recognition for speaker identification system.

Future work includes database of large number of speakers which can be tested using other classifier like SVM, GMM etc. More work is also needed in the area of feature extraction

REFERENCES

[1] B.S. Atal, “Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification,” Proc. J Acoustic Soc. Am., Vol. 55, No. 6, June 1974.

[2] Campbell, J.P. “Speaker recognition: a tutorial,” Proceedings of the IEEE Volume 85, Issue 9, Sept. 1997.

[3] M.S.Sinith, AnoopSalim, GowriSankar K, Sandeep Narayanan K V, Vishnu Soman, “A novel method for Text-Independent speaker identification using MFCC and GMM,” Proc. IEEE, Audio Language and Image Processing (ICALIP), Shanghai, Nov. 2010.

[4] Cholula, Puebla, “Speaker recognition using Mel frequency Cepstral Coefficients (MFCC) and Vector quantization (VQ) techniques,” Proc. IEEE, 978-1-4577-1326-2, Feb. 2012.

[5] Jayant M. Nayak, “Speaker Verification: A Tutorial,” Proc. IEEE, vol. 28, Issue 1, 0163-6804, Jan. 1990.

[6] M. A. Bashar, Md. Tofael Ahmed, Md. Syduzzaman, Pritam Jyoti Ray andA. Z. M. Touhidul Islam,” Text independent speaker identification using average pitch and formant analysis”, International Journal on Information Theory (IJIT), Vol.3, No.3, July 2014

[7] Noraziahtulhidayukamarudin ,S.A.R Al-Haddad ShaifulJahari-Hashim,Mohammad Ali Nematollahi,AbdRauf Bin Hassan,”Feature Extraction using Spectral Centroid and Mel Frequency Cepstral Coefficient for Quranic Accent Automatic Identification”,Research and Development (Scored), IEEE Student Conference on16-17 Dec. 2014

[8] M. Sahidullah and G. Saha, “Design, analysis and experimentalevaluation of block based transformation in MFCC computation forSpeaker Recognition,” Speech Communication, vol. 54, pp. 543-565, 2012.

[9] Jayanta Kumar Basu, Debnath Bhattacharyya, Tai-hoon Kim,”Use of artificial neural network in pattern recognition”, International Journal of Software Engineering and Its ApplicationsVol. 4, No. 2, April 2010

[10] WU Zunjing, CAO Zhigang,”Improved MFCC-Based Feature for Robust Speaker I identification”,Tsinghua Science and Technology Volume: 10, Issue: 2, April 2005, pp158-161.

[11] AbhilashaSukhwal, Mahendra Kumar,”Comparitive study of different classifiers based speaker recognition system using modified MFCC for noisy environment,” International conference on green computing and internet of things, 2016.

[12] Shafee, B.Anuradha,”Speaker identification and Spoken word recognition in noisy background using artificial neural networks,”ICEEOT, 2016.

[13] Ahmad R. Abu-El-Quran and Rafik **A.** Goubra,”Pitch-Based Feature Extraction for Audio Classification”,Haptic, Audio and Visual Environments and Their Applications, 2003. HAVE 2003. Proceedings. The 2nd IEEE International Workshop on,21-21 Sept. 2003.

[14] Mirghafori, N. and Heck, L.P, "An adaptive speaker verification system with speaker dependent a priori decision thresholds", Proc. ICSLP'02, pp. 589-592, 2002.

[15] Togneri, R., Pullella, D. (2011). “An Overview of Speaker Identification: Accuracy and Robustness Issues,”IEEE Circuits and Systems Magazine, Second Quarter 2011, 23-61.